Title: FIS Data Modeling Approach

Author: Susan Molina, South East Fisheries Science Center

Objective: We are receiving positive feedback from our business users on how we are implementing our Data Warehouse project, and specifically the way we are using dimensional modeling into our approach. This presentation will review at a high level:

- An overview of how we are using dimensional modeling and how we extended the method to address our specific needs
- The technical and business benefits we are achieving through our design approach
- Tips and techniques we have discovered along the way

High level overview of presentation content: The Southeast Fisheries Science Center (SEFSC) is currently holding design sessions to outline the approach for the Fisheries Information System ("FIS"). The FIS is intended to significantly extend the HMS Data Reconciliation System developed by the Science and Technology Office in Silver Spring, MD.

This objective is being accomplished by expanding the prior project to including more species and dealers, including data from the ACCSP (i.e., Atlantic Coastal Cooperative Statistical Program) data warehouse, trip ticket data from the states of North Carolina, South Carolina, Georgia, and Florida, and data from the GulfFIN repository. Furthermore, data from the Fisheries Logbook System will be quality assured against three sources to ensure that data that are provided by the states are adequately represented by the logbook collection effort.

The data shall be introduced into the SEFSC data warehouse following the pre-defined standards. This approach will align the data storage effort to work within the existing Dimensional Model of the SEFSC Data Warehouse. The process of extracting and loading the data will potentially produce exceptions such as 'orphan' records. Reports and maintenance tools will be developed to update orphan records to increase the quality of the data and produce more meaningful information.

The system is being designed to support

- Reconciliation of orphan and unknown records from different agency sources
- Maintenance tools to perform fuzzy logic against loaded data in an effort to improve data quality
- Creation of reconciliation reports and metadata repositories.

The team will follow the dimension model design methodology which provides several technical benefits such as

- Identifying conformed dimension which can be used among various attribute relationships
- Fact tables based on measured data and dimension surrogate keys, providing a many-to-many reporting relationship
- Orphan discovery during data loading, which identifies where data values are disconnected from dimension attributes. This process does not disrupt the overall warehouse data loading process.
- 'Unknown' record padding to support granular hierarchy levels

By developing the FIS data warehouse system as mentioned above, maintenance tools will be applied against the data to perform updates. This process will allow users to focus on the exception data, track down errors and omissions and raise the overall quality of the information accessed.

Furthermore, we will show how business users can access their information through the SEFSC dimensional model (schema) with simplicity and symmetry. Business users benefit from the simplicity because the data is easier to understand and navigate. The design is highly recognizable to business users, and easily extends to accommodate change.